CONTENTS

Anxiety as Related to Incongruencies Between Values and Feelings: *Paul McReynolds* 57

A Controlled Exploratory Investigation Into the Effects of Thorazine Upon Mental Test Scores of Chronic Hospitalized Schizophrenics: *John E. Nickols, Jr.* 67

Perspectives in Psychology: *Paul Swartz* 77

The Psychological Record, 1958, 8, 57-66.

ANXIETY AS RELATED TO INCONGRUENCIES BETWEEN VALUES AND FEELINGS[1]

PAUL McREYNOLDS

Veterans Administration, Palo Alto
and
Stanford University

Probably the most crucial problem in the understanding of the functional psychopathologies is the problem of anxiety. But despite the fact that anxiety is generally considered to be a significant factor in both normal and abnormal behaviors, our general understanding of its causes and correlates remains unsatisfactory. In a recent approach to this problem the author (8) proposed that anxiety may fruitfully be conceptualized as being a function of the quantity of unassimilated perceptual material. The purposes of the present paper are to summarize this position, to report briefly on some preliminary research to which it has led, and to describe a new technique being developed in connection with this research.

RATIONALE

In the course of living an individual undergoes many kinds of experiences, some anticipated and some unexpected, some sought and some due to the caprice of circumstance. It appears that such experiences, as realized phenomenally by the individual—as perceived by him, do not typically remain unorganized and in isolation, but rather tend to be incorporated into his overall apperceptive mass, i.e., into existent organized orderings of experience, which may be referred to as *perceptual systems*. The word *assimilation* can be used to refer to the process whereby an individual's percepts enter harmoniously and congruously into relation with other percepts, and are incorporated in a concordant manner into perceptual systems.

But on occasion an individual's percepts are not readily assimilable into previously organized experiencings, i.e., they do not "fit" and may actually be incongruent with previously integrated material. It is postulated that—until and unless appropriate perceptual restructuring occurs—percepts which are not assimilated will tend to accumulate, such that it is meaningful to speak of the quantity of unassimilated[2] perceptual data. It is hypothesized that anxiety is a *positive function of the magnitude of unassimilated percepts*, i.e., that the greater the quantity of unassimilated percepts the greater the degree of anxiety.[3]

This hypothesis could be tested most directly by testing the correlation between amount of unassimilated material and level of anxiety. Since there ap-

[1]From the Veterans Administration Hospital, Palo Alto. I am indebted to Anthony Fotenakes, James Bryan, Alice Beach, Mary Acker and Itsuye Sakai for their help in the collection and analysis of data, and to R. C. Tryon for his critical help in the preparation of this manuscript. A report of the studies reported here was read at the annual meeting of the California State Psychological Association, Los Angeles, 1957.

[2]Assimilation can possibly best be conceived of as occurring in various degrees, but for simplicity of presentation it is assumed that percepts either are, or are not assimilated.

[3]Probably a better formulation, but one which was not explored in the studies here reported, would be that anxiety is a function of the *ratio* of unassimilated material to the sum of assimilated plus unassimilated.

pears to be no direct way of measuring the amount of unassimilated material, however, it is necessary to utilize some indirect means of evaluating this quantity.

Probably the most important reason for percepts being unassimilable is the existence of *incongruencies* between them, or between the percepts involved and existent perceptual systems into which they would be assimilated.[4] By incongruency is meant characteristics of relationships between percepts or perceptual systems which make them difficult to reconcile or to harmonize, such as contradictions, inconsistencies, and anomalies. Two examples of incongruent material may help to make this point clear. If one has feelings of love and affection toward a given person, and yet this person behaves in a manner which arouses feelings of hostility, these new feelings might be somewhat difficult to assimilate. If one has high moral standards, and yet perceives that he has transgressed against these standards, these percepts might likewise be difficult to assimilate.[5]

Because of the implied relationship between incongruence and unassimilated material, it seemed reasonable to attempt to estimate the quantity of unassimilated percepts indirectly—by evaluating the degree of incongruency in the personality. This attempt would be based on the supposition that if an individual has marked incongruencies in the way he perceives different things and events and persons in life, then in the normal course of his behaviors he is likely to encounter a number of experiences which he cannot assimilate, and so will develop a large amount of unassimilated material. Hence, the above hypothesis can be restated to read: that anxiety is a positive function of incongruencies among one's perceptions and the perceptual systems into which they would be assimilated. It is to be noted that it is *not* held that anxiety is a function of personality incongruency *per se;* rather, the extent of incongruency is to be used to estimate the level of unassimilated material. Nor is it maintained that *all* incongruencies among perceptual systems necessarily lead to unassimilated percepts; on the contrary, individuals presumably sometimes avoid percepts in certain experiential areas in which there is marked incongruency, and so do not build up a backlog of unassimilated material despite the incongruency. Accordingly, it can be held only that a measure of incongruency should furnish a very rough measure of the level of unassimilated material, and hence of anxiety.

It is, of course, not possible—or in any event not feasible—to attempt to measure personality incongruency *generally.* One can hope only to measure certain kinds of incongruency. It was decided that a practical, first-attempt goal would be to measure incongruencies between *values* and *feelings.* The specific purpose of the studies to be described, then, was to test the hypothesis that degree of anxiety is positively related to extent of incongruency between values and feelings.

The relation between difficulty in assimilation, on the one hand, and incongruency between values and feelings, on the other, may now be discussed in more detail. Suppose that an individual has the experience of "bossing someone," and that he perceives this experience both with regard to *feelings* (whether he likes or dislikes it) and *values* (whether he considers it good or bad). We are interested in how he assimilates his feelings and his evaluations of this exper-

[4]Other factors which might contribute to difficulty in assimilation are: an excessively high rate of input of new percepts, a very marked degree of novelty, or variety, or change implied by the precepts, and an involvement with fears or uncertainties concerned with the *future,* such that they cannot be assimilated *now.*

[5]The concept of incongruency appears to be related to Festinger's (3) notion of "dissonance," to Peak's (9) concept of "disparity," and to concepts of ambivalence and mental conflicts generally—since the latter terms refer to various incongruencies within the personality.

ience. In order to illustrate the view being proposed, let us postulate four small perceptual systems, one each for the reception and incorporation of the percepts of Good, Bad, Like, and Dislike. So long as "bossing someone" is perceived simply as good, it can readily be assimilated into the 'Good' perceptual system; when perceived solely as bad it can readily be assimilated into that perceptual system; and so on for like and dislike. When something is perceived solely as good, bad, liked, or disliked, then, the situation as to assimilation would be rather simple.

We can suppose that small perceptual systems tend to become organized into larger, higher-order systems, that systems with a single dimension tend to be organized into larger systems with several dimensions. Such development of perceptual systems would tend to reduce overall complexity and would represent the integration of separate perceptual systems into larger systems (8, p. 296). In the present instance, it is reasonable to suppose that systems of Good and Like would tend to be combined into a single system, and Bad and Dislike into another. This supposition is based on the fact that the former concepts imply positive, approved characteristics, and the latter concepts negative and disapproved characteristics. More formally, it is assumed that individuals, in increasing the degree of assimilation of their percepts, tend to integrate separate, specific perceptual systems into higher-order systems. This assumption, as it relates to perceptual systems, is merely an extension of a general assumption that persons tend to assimilate percepts, i.e., that just as percepts tend to be assimilated into perceptual systems, so smaller perceptual systems tend to be integrated into larger ones.

If "bossing someone" is perceived as both good and liked, then, it could readily be assimilated into the larger perceptual system 'Good-Like'; if it is perceived as both bad and disliked, it could be assimilated into the larger perceptual system 'Bad-Dislike'; but if it is perceived as either good and disliked, or bad and liked, it could not be assimilated into either of the larger systems. In other words, difficulties in assimilation are likely to arise when one's percepts are incongruent with respect to the systems in terms of which they tend to be classified. The illustrative example just given, as well as the studies to be reported here, are based on the assumption that adults in our culture develop perceptual systems in which percepts of Good and Like, and Bad and Dislike, tend to be integrated into higher-order perceptual systems.

EXPERIMENTAL APPLICATIONS

Three studies directed to the hypothesis that anxiety is positively related to extent of incongruency between values and feelings will be described. These studies are not presented as being crucial experiments for the present theory. Probably their results could be rationalized in terms of conflict theory or of modern phenomenology. However, the experiments were developed in terms of the present theory, and are directly pertinent to it. They are best described as preliminary attempts to introduce empirical quantification into, and to develop specialized techniques for the study of, the proposed assimilation theory of anxiety.

Study 1

Subjects. — Twenty male NP patients on an admissions ward at the Palo Alto VA Hospital served as Ss. Including various diagnoses, they were selected by the ward staff as being capable of dealing successfully with the tasks used in the study.

Materials and Procedures. — In order to obtain estimates of how anxious the Ss appeared, standardized interviews were conducted—by a person not otherwise involved in the study. The interviewer rated the Ss' apparent anxiety on an 8-point scale. In order to obtain indications of how 'anxious' the Ss felt, each S was asked to rate himself on three 8-point scales, with regard to (a) how anxious he felt "right now," (b) how anxious he had felt "generally today," and (c) how anxious he felt "now as compared with the least and most anxious" he had ever felt. The last of the self-rating scales, which can be referred to as the self-comparison scale, was utilized because it was believed that it would provide an S with more meaningful end-points than the other two self-rating scales.

A special technique was devised in order to obtain an estimate of extent of incongruency between values and feelings. A box, 20 x 20 x 4 in., was constructed of wood, and the top lettered as indicated in Figure 1. Each of the 16 squares in the grid had a slot cut into it through which small cards could be dropped into the cubicles below. Fifty-six statements, such as "marriage," "reading the daily paper," "obeying my parents," "being an adult," "carrying a grudge," "my father," "myself," "to fight or quarrel," "relaxing," and the like, were typed on small cards.[6] Forty-six of these statements were selected as referring to the kinds of experiences, events, or relationships about which incongruencies might be expected to exist, and 10 were selected as referring to more innocuous matters.

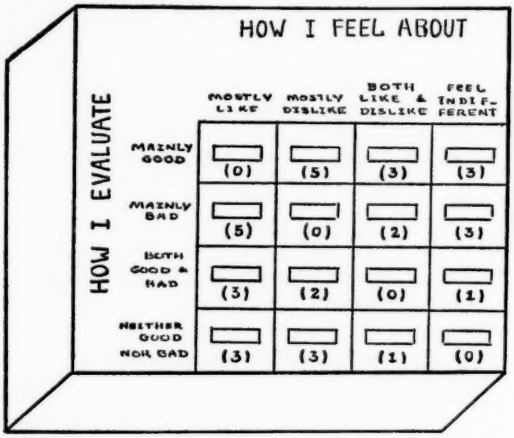

Fig. 1. Layout of box used to measure incongruency.

The S was given a tray containing the 56 cards in a predetermined random order. His task was to read the statement on each card and then, *first*, to decide how he *evaluated* the statement on the card—whether he considered it Mainly Good, Mainly Bad, Both Good and Bad, or Neither Good nor Bad, and *secondly* to decide how he *felt about* the statement—whether he considered his feeling to be Mostly Like, Mostly Dislike, Both Like and Dislike, or Feel Indifferent. On the basis of these two decisions, which determined both the appropriate row and the appropriate column, the S was to drop the card through the indicated slot in the top of the box.

[6]Mimeographed copies of the test items, rating scales, and directions used in all three studies reported here, as well as more recently devised materials, are available on request from the author.

It was reasoned that incongruencies between values and feelings would be indicated by the slots chosen. Thus, the diagonal squares—those labeled 0 in Figure 1—were assumed to reflect congruency between values and feelings. The two squares labeled 5 were assumed to indicate greatest incongruency. Incongruency values assigned to each of the other squares are indicated in Figure 1. All these assignments of values were done in advance on a judgmental basis. The values indicated in Figure 1 of course did not appear on the box used for testing. The *incongruency score* for each S was the sum of the values indicated by the squares into which he placed cards, the possible score range thus being from 0 to 280.

All Ss were seen individually. The interview was conducted first, then the S filled out the self-rating scales, and then the incongruency procedure was conducted.

Results. — Pearsonian correlation coefficients between incongruency scores and anxiety estimates were as follows: between incongruency and interviewer rating, .10; between incongruency and self-rating "right now," .14; between incongruency and self-rating "today generally," .41; and between incongruency and self-rating "comparison," .44. All of the coefficients are in the predicted direction, and the last two are significant at the .05 level on a one-tailed test. These results, though not confirmatory, tend to support the experimental hypothesis. It is to be noted that the correlations are highest for those anxiety estimates based on how the Ss themselves reported feeling. Intercorrelations between the different estimates of anxiety were: between interviewer and self-rating "right now," .32; between interviewer and self-rating "today generally," .21; between interviewer and self-rating "comparison," .02; and for the various self-ratings: "right now" and "today generally," .66; "right now" and "comparison," .24; and "today generally" and "comparison," .62.

Study 2

This study was undertaken in an attempt to extend the findings of Study 1, and in particular to try to develop a group test form of the incongruency technique.

Subjects. — The test procedures were group administered to 70 undergraduate college students, including 34 men and 36 women. During the course of administration it became apparent from Ss' questions that many of them did not clearly understand the instructions for one part of the incongruency test. The Ss were therefore asked to indicate, on their test forms, whether they felt they had clearly understood, or had not clearly understood the instructions. Forty-two Ss indicated that they did not clearly understand the directions, and accordingly their test forms were considered invalid. This left an effective sample of 28 Ss (20 men and 8 women).

Methods and Procedures. — Level of anxiety was estimated by the Taylor Manifest Anxiety Scale (12) and two self-rating scales. The two scales were not the same as those used in Study 1; rather, they represented attempts to improve upon those scales. One of these scales concerned how much stress and inner distress S had been "feeling recently," and the other was a "comparison" scale in which S was asked to indicate his present level of inner distress in terms of the worst and best he had ever felt. The word "anxiety" was not used in these scales, and the attempt was made to construct them at a level of sophistication appropriate to college students.

In order to estimate extent of incongruency between values and feelings two forms were prepared. One was called *Survey of Likes and Dislikes,* and included 100 items of the same type used in Study 1, each of which was to be marked as Mostly Like, Mostly Dislike, Both Like and Dislike, or Feel Indifferent. The other form was called *Evaluations Form,* and included 98 items of the same type, each of which was to be checked as Mainly Good, Mainly Bad, Both Good and Bad, or Neither Good nor Bad. Eighty-one items were common to both forms, and the incongruency score for each S was based upon the degree of incongruency between his indicated evaluations and feelings about the content of these 81 items. The Ss were not told that the two forms would be combined in this manner. The method of scoring was the same as in Study 1, except that squares given zero credit on Figure 1 were now given 1 point.

The tests were given in the following order: Survey of Likes and Dislikes, Taylor Anxiety Scale, Evaluations Form, and self-rating scales.

Results. — For the 28 Ss who indicated that they understood the instructions, and who could be considered as having taken the tests in a valid manner, the following correlation coefficients were obtained: between incongruency and Taylor MAS, .45; between incongruency and self-rating "recently," .47; and between incongruency and self-rating "comparison," .46. All of these coefficients are significant at the .01 level on a one-tailed test. Intercorrelations between the different estimates of anxiety were: between Taylor MAS and self-rating "recently," .40; between Taylor MAS and self-rating "comparison," .49; and between the two self-ratings, .49.

The fact that so many Ss failed to comprehend the test instructions is indicative of the inadequacy of the instructions, and represents a problem which must be overcome before the incongruency technique can become a useful experimental tool for group use. However, the data obtained upon the 28 Ss who reported that they understood the instructions clearly does not appear to have been biased in any way by the fact that other Ss did not understand clearly, and is thus acceptable as evidence with regard to the hypothesis being tested.[7]

College Board Scholastic Aptitude Test scores were available for the 28 Ss, and it was determined that the correlation between verbal aptitude and incongruency was .21, and between mathematical aptitude and incongruency .23. Neither of these coefficients are significant at the .05 level, and thus they suggest that incongruency scores are not a function of intelligence.

Study 3

In this study the same test materials and procedures were used as in Study 2, except that NP patients were used as Ss and were tested individually. The purpose of the study was not only to examine further the relation between incongruency and anxiety, but also to test the hypothesis that the mean of incongruency scores would be higher for the NP patients than for the students.

Subjects. — Forty-six male patients, mostly schizophrenics from continued-treatment wards, were seen for examination. Unlike the Ss in Study 1, these Ss were not selected in terms of their being probably capable of dealing adequately with the test materials. For various reasons, not all of the test measures were obtained on each S, so that N's for the various comparisons varied slightly.

[7] Intercorrelations between incongruency and anxiety for the 42 Ss who reported that they did not clearly understand the instructions were all insignificant, and were as follows: between incongruency and Taylor MAS, .18; between incongruency and self-rating "recently," —.01; and between incongruency and self-rating "comparison," .07.

Materials and Procedures. — As already noted, the same test materials were used as in Study 2. The order of testing in most instances was: Evaluations Form, Self-ratings, Survey of Likes and Dislikes, and Taylor MAS.

Results. — Coefficients of correlation between incongruency and the anxiety estimates were as follows: between incongruency and Taylor MAS ($N = 42$), .17; between incongruency and self-rating "recently" ($N = 45$), .08; and between incongruency and self-rating "comparison" ($N = 45$), —.11. None of the coefficients are supportive of the major hypothesis. Comparison of the students and patients on incongruency gave the following result. For 45 patients the mean incongruency score was 49.73, with $SD = 19.94$; for 28 students the mean was 41.18, with $SD = 17.95$. Comparison of the means yielded a t of 1.82, which with 71 df is significant at the .05 level on a one-tailed test. While the two groups differed on incongruency, the difference was not great, and there was considerable overlap.

With regard to the insignificant correlations between incongruency and anxiety, which are contrary to the findings in Studies 1 and 2, the implication is either that the major hypothesis is not valid, or that for some reason the measuring instruments did not furnish valid estimates in this study. With regard to the latter, it seems possible that the test instructions and the self-rating scales were worded at too sophisticated a level for the patients, even though all Ss reported that they understood clearly what they were to do. Another possibility is that some of the patients answered the Taylor MAS in a defensive and invalid manner. This possibility is rendered more reasonable by the following finding. MMPI records were available on 35 of the Ss, and since the Taylor MAS includes items from the MMPI it was reasoned that certain MMPI scores could be utilized to identify Ss who might have answered the Taylor MAS in an invalid manner. Accordingly, Ss were eliminated whose L, F, or K scores on the MMPI were over 70. This left 21 Ss. For these 21 Ss the correlation between incongruency and Taylor MAS was .54, which is significant at the .01 level on a one-tailed test.

DISCUSSION

The studies described above, viewed in their totality, lend at least tentative support to the hypothesis of a positive relationship between anxiety and extent of incongruency between values and feelings. However, in view of the experimental difficulty encountered in Study 2—in clearly conveying instructions to the Ss, and in view of the largely negative results obtained in Study 3, it cannot be concluded that the existence of the predicted relationship has been unequivocally established. Rather, the studies reported here should be viewed as preliminary experiments, and the need for further research is indicated. Such research should focus not only on improved measures of incongruency, but also on improved measures of anxiety. The general technique used for measuring incongruency in the present studies appears to have considerable potential. Further exploration with this technique will be aimed at the problems of item selection, reliability and clarity of instructions.

We may examine further the relationship between anxiety and incongruency. The present theory holds that incongruencies of all kinds, provided they are not in areas in which the S avoids perceptions, are related to anxiety, whereas in the present studies we dealt only with incongruencies between values and feelings, and these only within a rather limited range. This selectivity was necessary for reasons of experimental expediency, but if the theory is valid, then

an examination of the literature might reveal other studies which can be reinterpreted within the context of the present view.

For example, in a study of religious attitudes and manifest anxiety in college students, Funk (5) found a correlation of .43 between anxiety and "religious conflict." Since conflictive material can be interpreted as incongruent and unassimilated, this finding clearly is in accord with the present theoretical view.

Also pertinent to the present conceptualization of incongruency are studies which involve *Ss* using Q sorts in order to rate themselves, first, as to the way they see themselves and second, as to the way they would like to be. The correlation between these two sortings, often said to reflect degree of self-esteem or self-satisfaction, is directly a measure of a particular kind of "congruency," and hence, according to the present view, should be negatively related to level of anxiety. In a study by Block and Thomas (1) there is reported a correlation of —.69 between this particular measure of congruency and the Pt scale of the MMPI. Inasmuch as Pt is highly correlated with the Taylor A-scale (2), this amounts to a substantiation of the present interpretation.

Q-sort correlations between self and ideal ratings have been most used in studies conducted under the aegis of Carl Rogers. Rogers (10, 11) conceives of the degree of incongruency between self and ideal ratings—as reflected by low intercorrelations—as indicative of the degree of maladjustment. This interpretation, however, has led to difficulties (1, 6), one of the reasons therefor being that there is evidence (4, 7) that paranoid schizophrenics do not have a great deal more incongruency than do normals.[8] Since it cannot be maintained that paranoid schizophrenics are well adjusted, it seems doubtful that this particular measure of incongruency—between self and ideal ratings—can be defended as a measure of maladjustment. It appears to me, however, that such a measure can be interpreted, even for paranoid schizophrenics, in accordance with the view that it reflects to some extent the level of unassimilated material, and hence of anxiety. This interpretation would imply that paranoid schizophrenics are not necessarily notably more anxious than normals, but it would not imply that they are almost as well adjusted as normals. One can suppose that paranoid schizophrenics have constructed peculiar conceptual systems which, though consensually invalid, permit them to assimilate material which for them would otherwise be unassimilable, thus keeping anxiety relatively low. And, of course, it is just this peculiar conceptual schema which connotes the paranoid condition.

Friedman (4) did find evidence for what in the present context would be termed marked incongruency in paranoid schizophrenics when he obtained the correlations between their consciously stated self concepts and their inferred self concepts based on TAT stories. These correlations, as compared with normal *Ss*, were very low. It is to be noted, however, that this kind of incongruency, though it would imply abnormality, would not, according to the present theoretical view, necessarily imply anxiety. This is because the incongruency is based upon data which the *S* avoids the preception of, and so he would not build up a mass of unassimilated material. As pointed out earlier, incongruencies do not necessarily imply unassimilated material, but do so only to the extent that the individual actually obtains percepts in the areas of incongruency.

The present view of perceptual assimilation and incongruencies among percepts is in many ways quite similar to that of Rogers. Rogers (10, 11) has been interested in the extent to which incoming perceptual experiences are

[8] Friedman (4) reported correlations between self and ideal ratings of .63, .03, and .43 for normals, neurotics, and paranoid schizophrenics respectively.

assimilated into an individual's concept of himself, and he conceives of percepts inconsistent with the self picture as posing threat and creating tension. Anxiety is related to the individual's awareness of a discrepancy between his needs and his self picture. Rogers is consequently concerned with incongruencies between self and ideal perceptions, which is indicative of degree of maladjustment. The present view is that all kinds of incongruency among the variety of things one perceives—not just between self and ideal—help to determine the level of unassimilated material, and that this level determines the extent of anxiety, whether or not the *S* is aware of the incongruencies *per se*. Further, it appears that we must keep quite separate the concepts of anxiety and maladjustment.

SUMMARY

It is proposed that anxiety is a function of the magnitude of perceptual material which an individual has not assimilated. It is further proposed that a major reason for percepts being unassimilated is the degree to which they are incongruent with each other, or with perceptual systems into which they would be assimilated. It follows that degree of anxiety should be positively related to the extent of incongruency within the personality in those areas in which an individual does not avoid perceiving. Since it is not practicable to attempt to measure overall personality incongruency, it is necessary, in order to evaluate this hypothesis experimentally, to choose a specific kind of incongruency. This paper describes three studies designed to test the specific hypothesis that anxiety is related to incongruency between values and feelings.

The *Ss* were asked to indicate for each of a number of items what their feelings (Like, Dislike) were and what their evaluations (Good, Bad) were. The extent of discrepancies between their answers was taken as a measure of incongruency. Anxiety was estimated by interviewer ratings, by self-ratings, and by the Taylor Manifest Anxiety Scale. Two of the studies used psychiatric patients as *Ss*, and the other used college students.

Viewed in their totality, the findings are interpreted as tending to support the hypothesis, though not in an unequivocal and conclusive manner.

REFERENCES

1. BLOCK, J., & THOMAS, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, 51, 254-259.
2. BRACKBILL, G., & LITTLE, K. MMPI correlates of the Taylor Scale of Manifest Anxiety. *J. consult. Psychol.*, 1954, 18, 433-436.
3. FESTINGER, L., RIECKEN, H., & SCHACHTER, S. *When prophecy fails.* Minneapolis: University of Minnesota Press, 1956.
4. FRIEDMAN, I. Phenomenal, ideal, and projected conceptions of self. *J. abnorm. soc. Psychol.*, 1955, 51, 611-615.
5. FUNK, R. Religious attitudes and manifest anxiety in a college population. *Amer. Psychologist*, 1956, 11, 375 (Abstract).
6. HALL, C., & LINDZEY, G. *Theories of personality.* New York: John Wiley and Sons, 1957. Pp. 467-503.
7. HILLSON, J., & WORCHEL, P. Self concept and defensive behavior in the maladjusted. *J. consult. Psychol.*, 1957, 21, 83-88.
8. McREYNOLDS, P. A restricted conceptualization of human anxiety and motivation. *Psychol. Rep.*, 1956, 2, 293-312.

9. PEAK, H. Attitude and motivation. In M. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: University of Nebraska Press, 1955. Pp. 149-189.

10. ROGERS, C. *Client-centered therapy.* Boston: Houghton-Mifflin, 1951.

11. ROGERS, C., and DYMOND, R. (Eds.) *Psychotherapy and personality change.* Chicago: University of Chicago Press, 1954.

12. TAYLOR, J. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.,* 1953, 48, 285-290.

The Psychological Record, 1958, 8, 67-76.

A CONTROLLED EXPLORATORY INVESTIGATION INTO THE EFFECTS OF THORAZINE UPON MENTAL TEST SCORES OF CHRONIC HOSPITALIZED SCHIZOPHRENICS[1]

JOHN E. NICKOLS, JR.

The Hertzler Clinic

This project was designed primarily for the purpose of applying rigid scientific procedures to the task of investigating the basic effects of Thorazine upon mental efficiency. The problem area, mental efficiency, has been identified by Castner, Covington and Nickols (1). They proposed that Thorazine should be capable of lessening the disrupting effects of emotional disturbance, which in turn allows the *S* greater efficiency in setting his capacity to the task of dealing with immediate reality, particularly in relation to the less strictly intellectual functions. However, their conclusion was based upon the results of one small sample study that made no use of a control group.

The hypothesis is that a study of placebo versus Thorazine treatments will yield no significant differences in mean gains of mental efficiency for hospitalized chronic schizophrenic patients. The independent variable can be defined as Thorazine treatments and the dependent variable as mental efficiency. Because mental efficiency is not readily sampled in terms of absolute measures, gains and losses in test scores between successive testings were used to estimate change in mental efficiency.[2]

SUBJECTS AND PROCEDURES

Initially, the *Ss* consisted of an accidental sample of 40 chronic, female schizophrenic patients of an Iowa state hospital. Chronic wards were visited unsystematically and the first 40 suitable patients coming to the psychologist's attention were selected. Among other criteria of selection, all patients were believed

[1]This research was conducted at the Mental Health Institute, Independence, Iowa, as a special-interest project under the approval and sponsorship of the Iowa State Board of Control, particularly through the efforts of Board Member, Robert C. Lappan, and Psychology Director-Advisor, Lowell Schenke. The author is deeply indebted to his dissertation committee of Denver University: Lawrence Miller, Ph. D., Wilber Miller, Ph. D., Harry Moore, Ph. D., and particularly Stuart Boyd, Ph. D., who served as faculty advisor, and Joel Green, Ph. D., who served as statistical advisor. He also wishes to acknowledge consultation conferences with Leigh Peck, Ph. D., University of Texas, and Arthur Benton, Ph. D., and John Martire, Ph. D., University of Iowa.

The author wishes to express his thanks to the many employees of the State of Iowa who contributed to the progress of this research. J. O. Cromwell, M. D., Superintendent, approved and backed the research at the hospital level. Special acknowledgment is given to James Thomas for all administrations of the evaluative tests, his assistance in scoring the tests and assistance in supervising personnel, and for directing much of the follow-up study in the author's absence. Dr. J. Pietrzak and Dr. Henry Koseroski contributed their medical services. Substantial contributions were made by the Director of Nursing, Mrs. E. Kay, her female supervisory and psychiatric aide staff, and by J. Meuli, Coordinator of Special Activities, and his staff. Acknowledgment is also given to practicum student, Loyd Johnston, who administered the screening tests and assisted in the statistical analyses.

The author also wishes to thank the Smith, Kline and French Laboratories, which furnished all drugs, placebos and translations of foreign publications used for this project, and the secretarial staff of the Hertzler Clinic, Halstead, Kansas, which contributed typing services.

[2]The nature of "mental efficiency" is discussed elsewhere (3). In brief, it is used as a broader concept than "intellectual efficiency," allows for lowered reliability of most tests when used with chronic schizophrenics, infers that measurement is stabilized upon some relatively fixed and pertinent standard denominator, and can be used clinically to imply the existence of a relationship between the individual's present performance level and his actual mental capacity in coping with any task.

to have been free from Thorazine treatments for at least six months. Distributions of age and durations of illness for the sample do not deviate significantly from those for resident schizophrenics of the hospital or total population in Iowa state hospitals during 1951 to 1954.

Two groups of 20 Ss were matched on a group basis for age and for initial raw scores on the Graham-Kendall Test and Ammons Full-Range Picture Vocabulary Test. Attempts were also made to match the two groups on (a) complicating medical diagnoses, such as diabetes and syphilis, (b) lobotomized patients, and (c) recent somatic treatments. Statistical analysis revealed satisfactory comparisons on these and other variables. Other variables included (a) habitual level of psychomotor activity, (b) duration of hospitalization, (c) previous somatic treatment, (d) diagnoses, (e) intelligence quotients, and (f) time residing on the research ward prior to evaluative testing.

Ss were divided into four sections, with each including relatively well matched sub-groups of five control and five experimental Ss. Each section was bedded at random in one of four partially partitioned dormitories of the research ward.

The research ward was one of two newly and identically constructed wards to house chronic female patients. The two closed wards shared in common a large sitting and activities room.

Ward conditions were highly controlled so as to approach a laboratory situation as nearly as possible in a state hospital setting. Two psychiatric aides, especially selected for the project, were given specific mimeographed instructions. Privileges and activities were kept to a minimum, and those extended to any one S were also allowed to all 10 patients of her section. Activities included (a) some work on the ward, (b) television viewing and pastime activities in the common lounge, (c) short group walks and trips to the Canteen, (d) weekly dances and movies, and (e) up to two-hour week-end visits with relatives on the hospital grounds. Each S did not take up all activities.

Treatment consisted of a double-blind-treatment method. Ward personnel did not realize that placebos were used until after all research was terminated. They were told that the project was intended to test the effects of two drugs. Tablets had identical markings, whereas almost all Thorazine ampules were marked with a red dot and each placebo ampule had a blue ring. Each S's individualized drug supply was marked only with that patient's name, but personnel soon speculated that one drug was Thorazine and the other a different tranquilizer.

For each section, separate treatment schedules were placed on the ward. Ss received injections for the first week, then were switched to oral dosages. Daily dosages ranged up to, and included, a maximum of 400 mg. for each S. Total dosages were approximately 8,000 mg. during the first four to five weeks and 15,900 mg. for the full nine week program.

Gains in mental efficiency were measured primarily by weighted scores on the Wechsler-Bellevue Scales. These scales were administered before treatment, after four to five weeks, and after nine weeks of treatment (see Tables 2 and 3). Additional criteria tests, the Arthur Point Stencil Design Test, Form II[3], and Rorschach Test, were usually administered the afternoon of the same day on which any one S took the Wechsler-Bellevue Scales. Criterion testing was controlled for (a) examiner, (b) testing procedures, (c) fatigue, (d) physical environment,

[3]The author could find no standardized time credits for this test. Time credits were calculated by use of the probable errors for time credits obtained by the 40 Ss on initial testing.

(e) immediate effects of Thorazine, (f) time and sequence of testing, and (g) familiarity with testing and test materials. Scorings were re-checked at least twice, using at least two qualified scorers.

A screening battery of tests was administered by a practicum student. It consisted of (a) the Draw-A-Person Test, (b) Five Blot Test[4], (c) Graham-Kendall Memory-For-Designs Test, (d) Ammons Full-Range Picture Vocabulary Test, and (e) Facial Expression Pictures Test.[5] The battery was used for (a) screening, eliminating three untestable and three very deteriorated Ss, (b) matching, (c) controlling testing variables, and (d) speculations as to the effects of Thorazine treatments. Most of these tests were re-administered one to three days prior to re-testing with the evaluative criteria.

RESULTS

Thirty-six Ss completed the first four to five weeks of the study and 32 Ss completed the full nine weeks of treatment. One psychologist, a practicum student in psychology and two psychiatric aides were asked independently to rate patients according to specific standards of behavioral and psychiatric change during treatment.

TABLE 1

RESULTS OF PERSONNEL RATINGS FOR THE SYSTEMATIC STUDY*

	Average Percentages of Ss Responding to Treatment**						Percentage of Agreement Between Judges ***			
	Outstanding Improvement		Improved		Worse		Only for Ss Showing Change		For All Ss	
	C	E	C	E	C	E	C	E	C	E
Psychological Ratings	0	0	27	25	3	7	6	17	56	77
Psychiatric Aides	17	17	54	56	0	0	42	33	72	67

C = Control Group E = Experimental Group

* None of the differences are statistically significant, as evaluated by small sample *t*-ratios for proportions. All averaged percentage values of 25 or more are significant from zero, to the five per cent level of confidence, and values of 42 or more significant to the one per cent level of confidence.

** Each of these figures is based upon percentages of total patients and is expressed as an overall average for two judges who rated each *S* on psychiatric and behavioral improvement. Ratings for the first four to five and full nine week treatments were almost identical for any one judge.

*** These columns express the extent to which judges of each set agreed upon psychiatric and behavioral responses of individual Ss. Ratings for the first four to five and full nine week treatments were almost identical for any one judge.

[4]This test was specially compiled for this project by the author. It consists of five acromatic photographs of paint blots and was administered like the performance part of the Rorschach, to serve as a familiarizing experience for the Rorschach.

[5]This test was devised by the author, who has been collecting unanalyzed data for five years, and interpretations are made strictly from the more theoretical clinical frame-of-reference. The test consists of twelve acromatic photographs of paint blots, vaguely touched up to resemble human or animal faces. For this research, each *S* was asked if each card (a) resembled most closely animal, human, plant or object, and (b) appeared pleasant, unpleasant, a mixture, or neither.

Table 1 shows low ratings of improvement and very low agreement between judges for the psychological ratings of Ss improved. These raters were very cautious and observed the Ss for relatively short and scattered periods of time. The two psychiatric aides, who spent 40 hours per week on the research ward, show average judgments of approximately 50% improvement, including 17% outstanding improvement, for each group. For each group, ward personnel agree on at least two-thirds of their ratings and that at least one-third of each group improved. These ratings suggest that enough improvement occurred to allow psychological testing to sample the change, but they also suggest little differences between the groups.

TABLE 2

RESULTS ON THE EVALUATIVE CRITERIA FOR VARIOUS DURATIONS AND PROGRAMS OF THORAZINE TREATMENTS (*t*-ratios of 1.00 or less are indicated by "x")

Program	Mean Gains		Mean Gain Differences	*t* for Mean Gains		*t* for Uncorrelated Mean Gain Differences	Confidence Levels for Differences
	C	E	E − C	C	E	E − C	E − C
Wechsler-Bellevue Verbal Scale Gains (weighted scores)							
i2	3.3	1.5	−1.8	2.9	1.0	x	
i3	1.6	1.9	0.3	x	1.1	x	
i4	4.9	2.3	−2.7	2.3	x	x	
Wechsler-Bellevue Performance Scale Gains (weighted scores)							
i2	2.1	4.2	2.2	1.5	3.6	1.2	.24
i3	4.4	7.4	3.0	4.7	4.8	1.7	.10
i4	5.4	8.5	3.1	2.9	4.8	1.2	.24
Wechsler-Bellevue Full Scale Gains (weighted scores)							
i2	5.3	5.7	0.4	2.6	3.0	x	
i3	6.0	9.3	3.3	2.4	3.3	x	
i4	10.3	10.8	0.4	3.6	2.8	x	

C = Control Group E = Experimental Group

i2: $N = 18$ Results for the first four to five weeks of systematic treatments. Form II used on post-testing while Ss were on maintenance dosages of 100 mg. daily.

i3: $N = 16$ Results for nine weeks of systematic treatments. Form I repeated after one week of no treatment.

i4: $N = 12$ Results for nine weeks of systematic treatments plus 12 weeks follow-up treatments and activities. Form II repeated for post-testing after Ss were off Thorazine for at least one week.

Results of evaluations with standardized tests are presented in Tables 2 and 3. Table 2 shows that for the first four to five weeks of treatments the null hypotheses cannot be rejected significantly for any of the three Wechsler-Bellevue Scales. For the Verbal and Full Scales the results suggest that this situation is unaltered by a longer duration of treatment and re-testing without the immediate effects of Thorazine. Performance Scale results show that the Experimental Group had increased from a 2.2 to a 3.0 weighted score lead at the end of the

full nine weeks of treatments, and that the null hypothesis of no difference between the groups can be rejected at the 10 per cent level of confidence. One may speculate that this difference would reach significance, should it be possible to take into consideration the correlation due to matching groups.

TABLE 3

RESULTS OF TESTING WITH THE ARTHUR POINT STENCIL DESIGNS TEST, FORM II, BEFORE AND AFTER VARIOUS DURATIONS AND PROGRAMS OF THORAZINE TREATMENTS

(*t*-ratios of 1.00 or less are indicated by "x")

Program	Mean Gains		Mean Gain Differences	*t* for Mean Gains		*t* for Uncorrelated Mean Gain Differences	Probability for Mean Gain Differences
	C	E	E—C	C	E	E—C	E—C
Arthur Point Stencil Designs Test, Form II (raw scores)							
i2	1.4	0.3	—1.1	1.8	x	1.2	.24
i3	1.4	0.8	—0.6	2.2	1.7	x	
i4	1.4	0.0	—1.4	2.1	x	1.3	.21
Arthur Point Stencil Designs Test, Form II (with time credits)							
i2	8.9	2.3	—6.7	2.6	1.1	1.5	.15
i3	11.5	4.8	—6.8	2.7	1.6	1.3	.20
i4	10.6	2.8	—7.8	2.1	x	1.1	.28

C = Control Group E = Experimental Group

See Table 2 for interpretation of program symbols.

Results of the Arthur Point Stencil Designs Test show no significant differences between the groups, but the results consistently favor the Control Group for the first four to five and full nine week programs. Time credits, improvised for the test, served only to accentuate the Control Group's lead. This suggests that the probable action of Thorazine in facilitating improvement on Performance tasks is attributable to the nature of the tasks, or mental processes involved in coping with them, rather than to the time factor alone. It also suggests that appropriate use of the time factor with the more homogeneous tests should allow more effective use of item difficulty and an increase in precision of criteria testing.

EXTENDED STUDY

An attempt was made to estimate the stability of the results by the use of a total-push treatment program, using most of the same *Ss* of each group. The present study was extended for an additional 12 weeks. This program included Thorazine and placebo treatments, in conjunction with (a) industrial therapy, (b) extended privileges, including ground passes for up to 11 experimental and control *Ss*, and (c) twenty hours per week of activity therapy, including social, recreational, occupational, art and music therapies. Two psy-

chiatric aides had been asked to estimate daily dosages at which they had observed best behavioral responses while each S was participating in the systematic study. Dosages averaged approximately 200 mg. daily for each group.

Tables 2 and 3 indicate that results for the extended program are consistent with those for the first four to five weeks of the systematic study, and that no significant differences were found between the groups. The patterning of results for the three periods suggest that one should not overlook the possibility that a large sample replication study of increased experimental precision may show that Thorazine is better able to effect improvement of performance functions, but that the Control group may improve more on the Stencil designs and possibly on verbal functions. It is noteworthy that some mean gain differences for the total study (i4) are larger but less significant than those for the systematic study (i3); this may be due to the fact that the smaller sample did not allow the same statistical precision for the total study. The small test gains made during the longer and more intense follow-up study suggest the possibility that (a) repeated re-testing may have allowed the *Ss* to approach their improvement potential before participating in the total-push program, and/or (b) even though systematic and total-push treatment conditions are similarly capable of testing the effects of Thorazine, experimental precision should be increased by using the total-push program as an independent study to parallel systematic conditions.

RORSCHACH ANALYSIS

The Rorschach test was administered before treatment, after the systematic study, and after the follow-up, total-push program. For initial reaction time the Experimental Group shows significant decreases to acromatic, color, and all 10 cards for both treatment programs. Differences between the groups were significant for five of the six comparisons. These results suggest that Thorazine possesses the property of increasing spontaneity in chronic schizophrenics.

Small sample t-ratios for proportions were used to evaluate changes in 17 other scoring variables, for all *Ss* whose post-testing scores differed from their pre-testing scores. Probabilities for differences between the groups tend to distribute according to chance. That personality changes occurring during Thorazine treatments are likely to be highly individualistic is suggested from the observation that, while pre- and post-testing results differ markedly for the individual *S*, there is a general lack of consistency in direction and magnitude of change among the scoring categories for the group as a whole.

Directions of change on variables showing provisionally significant differences (5 to 10 per cent levels of confidence) tend to correspond to tentative conclusions made previously by the author (1). Using one-half values for additional responses, total numbers of responses tend to increase for experimental *Ss*, whereas sums for Y, color and popular responses tend to decrease. Although these trends were present for both the systematic and full 21 week treatment, only sum Y decreases were provisionally significant for both programs. These trends suggest that Thorazine helps lower the impact of emotional stress or threat and alleviates emotional disturbances or responsiveness, particularly in relation to external stimulation. FC decreases account for most of the decrease in color response, whereas Fc responses do not tend to change. This suggests that alleviation of emotional discomfort will not likely be accompanied by the development of more meaningful or satisfying affective or emotional relationships with social reality. On the contrary, the decrease in popular responses

suggests that conformity had become less important after treatment, and a general lack of change in human responses suggests no increase of interest in others. The increase in total number of responses suggests that the Ss experienced no loss of mental drive. No significant changes were found for approach and content.

ADDITIONAL COMPARISONS

Results of the systematic and follow-up programs of the present study were also compared with results of the Preliminary (Rusk) Study (1). This permits further speculation on (a) the stability of the results, and (b) identifying the possible weaknesses of the present study.

The Preliminary Study consisted of four and one-half weeks of Thorazine treatment and of total-push treatment efforts. Facilitating efforts included didactic group psychotherapy, extended privileges, recreation and social therapies, and some industrial and occupational therapy. In addition, 5 of the 10 Ss received two weeks of only recreational and social therapies and an additional five weeks of Thorazine treatment. No control group was used.

TABLE 4

A COMPARISON OF THE PRESENT STUDY WITH THE PRELIMINARY STUDY FOR RESULTS OF WECHSLER-BELLEVUE SCALES WITH VOCABULARY AND OBJECT ASSEMBLY TESTS OMITTED

(*t*-ratios of 1.00 or less are indicated by "x")

Program		Mean Weighted Score Differences		*t* for Mean Gain Uncorrelated Differences		Confidence Levels for Differences	
		R—C	R—E	R—C	R—E	R—C	R—E
Full Scale Weighted Score Gains	i2	10.7	11.7	3.3	3.9	.01	.01
	i3	9.9	8.0	3.0	2.0	.01	.06
	i4	7.6	6.9	1.9	1.6	.07	.13
Verbal Scale Weighted Score Gains	i2	1.3	2.8	x	1.2	—	.24
	i3	2.4	1.9	x	x	—	—
	i4	—0.1	1.7	x	x	—	—
Performance Scale Weighted Score Gains	i2	9.5	8.8	5.3	5.6	.01	.01
	i3	7.5	6.2	4.9	2.1	.01	.05
	i4	7.7	5.2	3.8	2.7	.01	.02

R: Results of the preliminary study, or Rusk Group. (No control group was used.)

C: Results of the control group of the present, or Independence study.

E: Results of the experimental group of the present, or Independence study.

See Table 2 for interpretation of program symbols.

These comparisons, presented in Table 4, show no significant differences between the Preliminary Group and either group of the present study for Wechsler-Bellevue Verbal Scale gains, whereas the Preliminary Group did significantly better than either Independence group on Performance Scale gains for all three programs. This suggests that at least the Performance Scale results of the present study may be duplicated in other Thorazine-centered programs. With the possible exceptions of the initial use of a total-push treatment and use of 12 sessions of didactic group psychotherapy, the less prolonged and no more intense preliminary program suggests that *S* variables account for its superior results over the present study on Performance and Full Scale gains. Probable *S* variables are that the Preliminary Study included substantially greater percentages of Ss who (a) had not previously received any form of tranquilizing drugs, (b) had been hospitalized for less than one year, (c) were active and overactive, psychomotorly, and (d) were acutely or emotionally disturbed. However, these variables may also indicate favorable prognoses for any psychiatric treatment.

Results for all individual tests and Wechsler-Bellevue sub-tests used in this study were inspected for statistical significance and consistencies of gains. Control Ss improved significantly more on the Full-Range Picture Vocabulary Test. They tended to improve substantially more on the Arthur Point Stencil Designs, Test Form II, and on the Arithmetic Test. Experimental Ss did significantly better on affect perception of the Facial Expression Pictures, or in attributing affectual meanings in common with 20 senior nursing students, and on the Object Assembly Test. They tended to improve substantially better on the Picture Arrangement Test and the Digit Symbol Test.

Pearson product-moment correlations were run between each of four *S* variables and the three Wechsler-Bellevue Scales, for the gains made during the systematic and systematic plus total-push treatments. Rho correlations were also computed with composite gain scores (Wechsler-Bellevue Full Scale plus Stencil Designs plus one-fourth Full Range minus one-fourth Graham-Kendall score gains).

Twenty-two comparisons were made for each of four *S* variables. The number of significant differences between the groups tend to distribute within chance expectancies, but certain precautions may be suggested for precise research. For 18 of the 22 comparisons age correlated negatively, and deterioration (initial Graham-Kendall scores) positively, with improvements in mental efficiency, but use of only Ss between 25 and 50 yr. tended to erase the age trend. Duration of hospitalization does not seem to be related to therapeutic improvement, except where durations are of less than four, particularly of less than two, years. Initial Full-Range score comparisons were inconsistent.

Five rank order correlations were computed to compare each initial standardized criterion test and each Rorschach variable with composite mental efficiency gains. No significant correlations or differences were found between therapeutic gains, on the one hand, and initial Stencil Designs scores, with or without time credits, initial Full Scale, Verbal or Performance I.Q.'s, on the other hand. For 25 Rorschach variables, significant differences tended to distribute according to chance, but future research might give principal consideration to the study of initial reaction time, sum Y responses, particularly C′ responses, and the color ratio. Other variables may include whole responses, FC responses, and possibly Dd and total number of responses.

For the reader who wishes to speculate on prognostic indicators, which may be of value in selecting Ss specifically for Thorazine treatments, Rho correlations between initial Rorschach variables and composite score gains suggest that better therapeutic responses may be associated with initially higher sum W, sum Y and sum C′ responses, and initially lower color-ratios.

SUMMARY AND DISCUSSION

This project was designed as a matched group investigation into the effects of Thorazine upon the mental efficiency of chronic, hospitalized schizophrenics. Sixteen experimental and control Ss completed nine weeks of treatment under laboratory-like conditions, and 12 Ss of each group completed an additional 12 weeks of total-push treatment conditions. Statistical analyses of mean gains do not allow significant rejection of the null hypotheses for Wechsler-Bellevue Verbal, Performance, or Full Scales Test and Arthur Point Stencil Designs Test. This suggests that Thorazine is incapable of producing clear-cut and significantly measurable effects on mental efficiency, for the Ss, experimental precision, and conditions employed in this study. Rorschach evaluations and personnel ratings also failed to show significant evidence that the effects of Thorazine differ substantially from those obtained with placebos under the same therapeutic conditions.

Several factors suggest that, in order to interpret safely the essentially negative character of the results, the implications of several trends in the results of the present study should be ruled out in future research, by use of large sample replication studies of increased experimental precision. Possible weaknesses of this study have been discussed in the text.

Inspection of trends of change among the results allow several speculations which should be followed up in future research. Thorazine treatments seem to be consistently associated with markedly improved efficiency in coping with performance tasks; but the drug seems to be ineffective, if not disadvantageous for improving performances to certain verbal or the more strictly intellectual functions. Projective test evaluations suggest that Thorazine may allow increased spontaneity and alleviate emotional responsiveness and experiences of stress reactions to external stimulation, which do not seem to be accompanied by improvement in the basic illness, mental content, or meaningful personal relationships with the environment.

The overall pattern of results suggests that the drug may effect favorable influences upon disorders that are due primarily to the disrupting effects of emotional disturbances upon mental efficiency and use of previously established habit-formations and behavioral control, particularly where the pathological reaction-sequence is secondary to and/or serves to alleviate emotional discomfort aroused by external stress-provoking stimuli, but impotent or possibly resistant to helping disorders that are due primarily to aberrations of mental content or the more strictly intellectual functions. These speculations suggest that future research should investigate the possibility that Thorazine is effective with basic emotional and social disorders, but not necessarily and possibly disadvantageous with basic mental and intellectual disorders. More specific speculations can be extracted from the text of this paper.

REFERENCES

1. CASTNER, C. W., COVINGTON, C. M., & NICKOLS, J. E., JR. The effects of a Thorazine-centered treatment program, with psychological evaluations. *Texas Rep. Biol. Med.*, in press.

2. COVINGTON, C. M., & NICKOLS, J. E., JR. Treatment of white female patients at the Rusk State Hospital from February 1 to October 1, 1954. Paper read at Texas & Mexico Neuropsychiat. Ass., Oct., 1954.

3. NICKOLS, J. E., JR. An exploratory investigation into the effects of Thorazine upon mental test scores of chronic, hospitalized schizophrenics. Unpublished doctoral dissertation, Univer. of Denver, 1957.

The Psychological Record, 1958, 8, 77-85.

PERSPECTIVES IN PSYCHOLOGY

VII. THE CRITERIA OF VALIDITY IN OBSERVATIONAL ANALYSIS[1]

PAUL SWARTZ

University of Wichita

> The main body of psychology started its career by putting the wrong foot forward and it has been out of step with the march of science much of the time since. Instead of beginning with studies of the whole person adjusting to a natural social environment, it began with studies of a segment of a person responding to a physical stimulus in an unnatural laboratory environment. Consequently, after a century of diligent application, psychologists still lack sufficient ordered knowledge of everyday social behavior. Their attempts to overcome this handicap by adding or associating in some way the psychological processes which have been scientifically investigated in the laboratory have not been notably successful.
>
> —*Assessment of Men.*

Several recent commentaries on the current situation in psychology have raised questions regarding the validity of the experimental approach to behavior. The suggestion is made that experimentation has been overstressed in psychological work and that the possibilty of making greater use of *observational* techniques should be carefully explored. In an especially frank and perceptive formulation of this argument, Paul Lazarsfeld (7) has expressed doubt that such complex processes as voting or buying behavior are amenable to experimental analysis. "You can experiment with how people bet," suggests Lazarsfeld. "But if you want to know how people vote, and how people buy, I don't think you can really experiment. You have to do systematic analytical observations."

The present writer has stated the issue in more general terms.

> Simple behaviors, activities in which physiological processes predominate, behaviors coordinated with minimal stimulation and stimulus deprivation—with these forms of action a valid experimental psychology seems not only possible but distinctly probable. But what of the more complex behaviors, in which conduct is coordinated with very complex patterns of stimulation and in which the factors of past behavior history and social conditioning are of paramount importance? What is the outlook here (12, p. 122)?

[1]Part of this paper was presented at the annual meeting of the Kansas Psychological Association in Wichita, Kansas, on April 19, 1958.

Tentatively, the position is taken that while "we must not give up experimenting with complex behaviors," provision should be made for the possibility that "the really significant 'break-throughs' " in understanding may result from "the perceptive observations of the non-experimentalist 'amateur' and non-experimentalist professional," rather than from controlled laboratory study.

Observational analyses of human behavior are apt to be regarded with scornful disapproval by the defenders of orthodoxy in psychology and are liable to the charge of anecdotalism. "There is an Idol of the Laboratory," writes Joseph Wood Krutch, "as well as of the Market Place" (5). Intolerant and unenlightened as this attitude may appear, there is considerable justification for it in the continuing, centuries-long struggle to establish psychology as a science. The observational approach is not susceptible of such precise definition or circumscription as the experimental approach. Neither is it as reliably self-corrective. It can easily degenerate into a fruitless, paralyzing rationalism. If the lessons of history are to be our guide, then the burden of proof clearly rests on those who propose the observational approach and not on their critics. "The disillusion with the laboratory" (4), must not be allowed to usher in the return of subjectivism or prelude the surrender of scientific method. The intent of the observational approach is to broaden the scope and significance of psychological inquiry, not to emasculate it.

The question which must be answered, and answered well, by the observational psychologist is the question of criteria. What standards exist for judging the validity of conclusions based on observational findings? Are the traditional criteria of prediction and control applicable in observational analysis? Or are there other standards to invoke? In the present paper we propose to make a preliminary examination of this problem. No solutions are offered; the problem admits of no easy solution. We have only a few suggestions to make, which we hope will help clarify the problem and bring certain aspects of it into sharper focus.

THE CRITERIA OF VALIDITY IN OBSERVATIONAL ANALYSIS

The observational approach in psychological work may assume several different forms. For our purposes we can distinguish (a) systematic analytical observation, as proposed by Lazarsfeld (7), (b) clinical observation, and (c) non-systematic, reflective observation, as embodied in the art of the novelist or playright and in the work of the professional psychologist who prefers a non-systematic, natural behavior-world type of study. Each of these three observational methods incorporates certain unique features. It is therefore necessary to consider them singly in discussing the question of criteria of validity.

Systematic Analytical Observation

The systematic analytical observation of behavior is well described by Lazarsfeld in a brief reference to some of his own work.

> Let me come back once more to the type of voting studies I am especially interested in. We take a sample of people in, say, Elmira, New York, and interview them every month. We ask them for whom they want to vote, and then see how they change from month to month. We have done that with car buyers; we have kept potential car buyers under observation for nine months and have tried to study the decision process

> by re-interviewing the same people in various stages of their approach to a final purchase.
>
> Many such decisions can best be studied by making repeated observations and seeing how the outcome at one point affects the turn of affairs in the next time period. We have data on disc jockeys . . . . How do they decide which record to play next? We made very detailed observation with people sitting beside those disc jockeys and watching how they pick the record, and then we interviewed those disc jockeys. We can tell what records they play next as a result of what telephone calls come in . . . (7, pp. 102-103).

Of the three forms of observational study that we have distinguished, it is obviously systematic analytical observation that most nearly resembles experimental research. By the careful selection of behavior samples and the skillful design and standardization of procedures, the controlled conditions of study characteristic of the laboratory may be significantly approximated by the psychologist who uses this method. Furthermore, systematic analytical observation lends itself to quantitative treatment. Lazarsfeld, for example, in connection with his own work, suggests the relevancy of time series analysis. "The models to be looked for," he writes, "have a similarity to business cycle analysis of the economists" (7, p. 103).

It should be readily apparent that the determination of the validity of conclusions about behavior derived from systematic analytical observation does not present a special problem. The criterion of prediction is clearly applicable. We have already cited one example of applying this standard in describing Lazarsfeld's disc jockey study. Successful prediction based on conclusions derived from observational analysis increases the likelihood of their being valid. Unsuccessful prediction casts doubt on their validity.

Clinical Observation

The second type of observational method that we have distinguished is clinical observation. Psychologists, psychoanalysts, and psychiatrists have all paid considerable attention to the question of validating knowledge derived in the clinical situation. The problems here are very serious and very complex. A good, representative introduction to the basic issues involved can be found in the 1950 Hixon Lectures on *Psychoanalysis as Science* (11). *Mutatis mutandis,* the observations made by the participants in this series can be applied to any attempt to validate clinical knowledge, regardless of the particular theoretical orientation of the clinical investigator.

The viewpoints expressed by the participants in the Hixon Lectures are sufficiently diverse and well-developed to provide a basis for our discussion. Hilgard (3), of course, upholds the possibility of controlled, laboratory validation of psychoanalytic concepts and techniques. Although he recognizes that "so many of the experiments give merely trivial illustrations of what psychoanalysts have demonstrated to their own satisfaction in clinical work," and that "experimental work thus far bears most directly only on the most superficial aspects of psychoanalytic theory, while many of its deeper problems are scarcely touched," nevertheless, he stoutly maintains that "if we are able to design experiments appropriate to the more superficial aspects, we can move on to deeper stages." He cautions, however, that "laboratory experimentation

that preserves anything like the richness of a psychoanalysis will be very difficult indeed, if not, perhaps, impossible."

Hilgard's optimism regarding an experimental approach to psychoanalysis is not shared by E. Pumpian-Mindlin, himself a psychoanalyst. In response to the question: "Are, then, psychoanalytic data and theory verifiable in the laboratory?" he writes:

> I am afraid that we must face the fact that they are not subject to direct verification in this manner at the present time. Derivative data, such as the mechanisms of defense or the presence of conflicts and of unconscious forces, may be verified, but the detailed derivation of these factors in the analytic sense at present cannot be subjected to any type of experimental validation, except by the use of the analytical method itself. I do not underestimate the type of experimentation and clinical validation offered by Dr. Hilgard in his lectures . . . but the extension of this material beyond these limits is not yet feasible (10, p. 155).

This does not mean, continues Pumpian-Mindlin, that the theory and data of psychoanalysis cannot at all be validated. While the results will not satisfy the stringent criteria employed in the biological and physical sciences, "it is possible," he writes,

> to obtain objective evidence of at least certain general hypotheses of psychoanalysis. Certainly it is possible to investigate in the laboratory the nature and quality of the underlying forces present in human beings, as to a certain extent has already been done (p. 155).

The subject matter of psychoanalysis, argues Pumpian-Mindlin, is "on a different integrative level" than that of the "exact sciences." The definiteness and specificity demanded of hypotheses and principles in these sciences cannot be expected in psychoanalysis. Psychoanalytic observation cannot be as highly controlled as investigation in the physical and biological disciplines. At present, it is concluded, psychoanalysis must be content with "establishing what appear to be significant, but not exclusive, correlations rather than specific causal relationships."

To this writer the most cogent analysis presented in the Hixon Lectures is that formulated by Lawrence S. Kubie (6). His argument is brilliantly conceived and masterfully executed. No review can hope to do it justice. It must be read in toto. All that we will attempt to do in this paper is to underscore a few of its major points.

The experimental scientist, writes Dr. Kubie, does not at this time know enough about psychoanalysis to determine what questions are appropriate and significant. The first task of the experimentalist, he observes, is to:

> make himself thoroughly familiar with phenomena as these occur in nature, ascertaining what can be proved with the unaided eye and ear before deciding what to subject to experimental verification. Otherwise, investigators may use complex methods to prove something which needs no proving, precisely because it is on the surface for all to observe, either during the naive phases of childhood or in the facts of illness which are familiar to the clinician (p. 64).

Laboratory validation of such data, continues Kubie, is redundant, having mainly an educational value. Of course, he adds, the precision and refinement of data which the laboratory makes possible is not superfluous.

> My argument is solely against the uselessness of making pallid facsimiles in the laboratory of data which are already manifest in nature, merely to get around the human reluctance to look human nature in the eye. This comment is relevant to no small part of the experimental work that has been documented by Sears . . . , some of which was referred to by Professor Hilgard. Many of these laboratory charades are pedestrian and limited demonstrations of things which have been proved over and over again in real life . . . . Experimental facilities should not be wasted on issues which are already clearly proved, and to which human bias alone continues to blind us. The experimentalist should rather take up where the naturalist leaves off (pp. 64-65).

In this lucid and courageous exposition of one of the fundamental principles of scientific work, Kubie has pointed unerringly to the major criticism that the observational student of behavior makes of traditional experimentation in psychology, namely, its almost studied neglect of events in the natural behavior-world. How much of our so-called basic research in psychology pales beside the richness and variety of real life activity! How much is but the controlled investigation of laboratory fictions! Lazarsfeld has deplored this self-imposed limitation on psychological work in a brief, but penetrating discussion of the data of science. He writes:

> Experiments might be the only source of data for the physicist and they might be useful for the social scientist. But they are for us certainly not the main material we are working with today. Don't forget that we have data for something like 2,500 years. Plato has developed very interesting propositions on human behavior in society and, if we now come and develop models which somehow organize or clarify what Plato has to say, we are doing an important job. It isn't true that the only data in the social sciences are those which have been collected in the laboratories in the last 50 years (7, p. 99).

There is one apparent weakness in Kubie's analysis which the present writer readily acknowledges. How does one distinguish between data which require no experimental verification because they are "already manifest in nature," and data which appear to be but are not really manifest in nature? How, in other words, does the naturalist avoid the error of making "the world is flat" type of assertion?

The problem is not insoluble, although the solution is not fool-proof. We assume, first of all, that our naturalistic investigator, like the cultural anthropologist, is a well-trained student of behavior. Second, we assume that only those data concerning which a reasonable consensus can be reached will be accepted as "manifest in nature." Third, we assume that where the results of naturalistic observation contradict the already accepted body of knowledge, more complex methods of verification, not necessarily experimental in nature, will be applied. Will these safeguards eliminate all error? Certainly not. But no science has or ever will devise an error-free investigative technique. If there

is any possibility of advancing our understanding of human behavior by naturalistic observation, we should certainly be willing to bear the risks involved, recognizing, of course, that this is only one phase of the investigative operation. Hilgard expresses some appreciation of the usefulness of naturalistic study when he writes: "If we are wise enough to make the most of naturalistic observations, we may find some clues that will help us to design better experiments" (3, p. 35).

"The major contribution of experimental science," contends Kubie, "will not be limited to confirming psychoanalytic observations in the laboratory but rather to providing psychoanalysis with instruments of greater qualitative and quantitative precision." The first problem to be solved is how to record adequately the therapeutic process *"without at the same time distorting the process."* Kubie proposes an institute for research in psychoanalytic psychology, an organization which would coordinate the work of many disciplines: neurophysiology, biophysics, biochemistry, pharmacology, clinical psychology, cultural anthropology, biostatistics and, of course, psychoanalysis. He calls such an institute "a dream for the future."

In our limited judgment Kubie has correctly stated the relationship between clinical observation and controlled experimentation, and has argued convincingly for the validity of the former as a means of studying behavior. We would add only one step to his program, namely, a sociological analysis of the general and specific cultural circumstances in which the clinical knowledge is derived. Pumpian-Mindlin hints at the necessity for this when he calls attention to the fact that Freud "arose in a definite historical period and was subject to its influences." If we accept the relationist thesis of the sociology of knowledge, viz., "that it lies in the nature of certain assertions that they cannot be formulated absolutely, but only in terms of the perspective of a given situation," and if we assume that assertions about human behavior are mainly of this type, then it logically follows that the first step in determining the validity of psychological knowledge must be "an analysis based on the sociology of knowledge" (9).

Non-systematic, Reflective Observation

We come now to the third type of observational analysis, namely, non-systematic, reflective observation, as embodied in the art of the novelist or playright and in the work of the professional psychologist who prefers a non-systematic, natural behavior-world type of study. Here we are on dangerous ground. Judged by the standards of modern, experimental psychology, this approach to behavior is easily the *least* respectable of all the observational methods. Even though many psychologists would agree with Verplanck (13), that novelists and playrights "do a remarkably good job of giving plausible accounts of behavior, often in terms that seem pertinent," nevertheless, these same psychologists would probably acquiesce in the prevailing opinion that "literary psychology" is an alien rite in the House of Science. Allport's plea (1) for pluralism in the methods of psychology, as expressed in his belief that "whatever contributes to a knowledge of human nature, is an admissible method to science," seems not to have had much impact on traditional psychological thinking.

The psychologist's refusal to entertain seriously the possibility that literary creation can make a significant contribution to the understanding of human behavior rests, we believe, in what Karl Mannheim (9) would call "a one-sided and narrow orientation to the problem of knowledge." Consider the following observation, relative to the question of scientific vs. so-called pre-scientific

knowledge and understanding. It appears in Mannheim's classic work, *Ideology and Utopia*. Compare it to the typical "holier-than-thou" attitude of the experimental psychologist. We believe that the program suggested is an eminently sane and practical one, which might even have as a beneficial secondary effect, the re-establishment and strengthening of the psychologist's contact with real life behavior.

> If, however, it is true, that life affords possibilities of knowledge and understanding even where science plays no part, it is no solution to designate such knowledge as "pre-scientific" or to relegate it to the sphere of "intuition," simply in order to preserve the purity of an arbitrary definition of "science." On the contrary, it is above all our duty to inquire into the inner nature of these still unformulated types of knowledge and then to learn whether the horizons and conceptions of science cannot be so extended as to include these ostensibly pre-scientific areas of knowledge (p. 165).

The problem of validating knowledge derived from non-systematic, reflective observation can be best handled in two parts, corresponding to the two extreme instances of this form of knowledge. There is, first of all, knowledge derived from non-systematic, reflective observation that is confirmed in systematic analytical observation, clinical observation, or experimental analysis. This confirmation may be either intentional or accidental, witting or unwitting. Assuming that we can eventually agree on a set of criteria for the validity of clinically derived knowledge, such confirmation would seem to provide presumptive support for the validity of the non-systematically, reflectively derived knowledge.

One of the best examples of this manner of validating knowledge can be found in the frequent clinical confirmation of *ambivalence* as an important characteristic of affective behavior. Freud (2) attributes the coining of this term to Bleuler. Yet, but for the naming of the behavior, the ambivalent reaction was described with great insight over two thousand years ago by Lucretius (8), the celebrated Roman poet. In *Book IV* of his poem, *The Nature of the Universe*, in which Lucretius discusses sensation and sex, he writes:

> Do not think that by avoiding grand passions you are missing the delights of Venus. Rather, you are reaping such profits as carry with them no penalty. Rest assured that this pleasure is enjoyed in a purer form by the healthy than by the love-sick. Lovers' passion is storm-tossed, even in the moment of fruition, by waves of delusion and incertitude. They cannot make up their mind what to enjoy first with eye or hand. They clasp the object of their longing so tightly that the embrace is painful. They kiss so fiercely that teeth are driven into lips. All this because their pleasure is not pure, but they are goaded by an underlying impulse to hurt the thing, whatever it may be, that gives rise to these budding shoots of madness (pp. 163-164).

This example is but one of many that could be cited in which non-systematically, reflectively derived knowledge, through clinical or experimental confirmation, clearly makes a contribution to psychological understanding.

Two thousand years, we would all agree, is a long time to wait for clinical confirmation of a naturalistic observation. Surely, there must be some alterna-

that preserves anything like the richness of a psychoanalysis will be very difficult indeed, if not, perhaps, impossible."

Hilgard's optimism regarding an experimental approach to psychoanalysis is not shared by E. Pumpian-Mindlin, himself a psychoanalyst. In response to the question: "Are, then, psychoanalytic data and theory verifiable in the laboratory?" he writes:

> I am afraid that we must face the fact that they are not subject to direct verification in this manner at the present time. Derivative data, such as the mechanisms of defense or the presence of conflicts and of unconscious forces, may be verified, but the detailed derivation of these factors in the analytic sense at present cannot be subjected to any type of experimental validation, except by the use of the analytical method itself. I do not underestimate the type of experimentation and clinical validation offered by Dr. Hilgard in his lectures . . . but the extension of this material beyond these limits is not yet feasible (10, p. 155).

This does not mean, continues Pumpian-Mindlin, that the theory and data of psychoanalysis cannot at all be validated. While the results will not satisfy the stringent criteria employed in the biological and physical sciences, "it is possible," he writes,

> to obtain objective evidence of at least certain general hypotheses of psychoanalysis. Certainly it is possible to investigate in the laboratory the nature and quality of the underlying forces present in human beings, as to a certain extent has already been done (p. 155).

The subject matter of psychoanalysis, argues Pumpian-Mindlin, is "on a different integrative level" than that of the "exact sciences." The definiteness and specificity demanded of hypotheses and principles in these sciences cannot be expected in psychoanalysis. Psychoanalytic observation cannot be as highly controlled as investigation in the physical and biological disciplines. At present, it is concluded, psychoanalysis must be content with "establishing what appear to be significant, but not exclusive, correlations rather than specific causal relationships."

To this writer the most cogent analysis presented in the Hixon Lectures is that formulated by Lawrence S. Kubie (6). His argument is brilliantly conceived and masterfully executed. No review can hope to do it justice. It must be read in toto. All that we will attempt to do in this paper is to underscore a few of its major points.

The experimental scientist, writes Dr. Kubie, does not at this time know enough about psychoanalysis to determine what questions are appropriate and significant. The first task of the experimentalist, he observes, is to:

> make himself thoroughly familiar with phenomena as these occur in nature, ascertaining what can be proved with the unaided eye and ear before deciding what to subject to experimental verification. Otherwise, investigators may use complex methods to prove something which needs no proving, precisely because it is on the surface for all to observe, either during the naive phases of childhood or in the facts of illness which are familiar to the clinician (p. 64).

Laboratory validation of such data, continues Kubie, is redundant, having mainly an educational value. Of course, he adds, the precision and refinement of data which the laboratory makes possible is not superfluous.

> My argument is solely against the uselessness of making pallid facsimiles in the laboratory of data which are already manifest in nature, merely to get around the human reluctance to look human nature in the eye. This comment is relevant to no small part of the experimental work that has been documented by Sears . . . , some of which was referred to by Professor Hilgard. Many of these laboratory charades are pedestrian and limited demonstrations of things which have been proved over and over again in real life . . . . Experimental facilities should not be wasted on issues which are already clearly proved, and to which human bias alone continues to blind us. The experimentalist should rather take up where the naturalist leaves off (pp. 64-65).

In this lucid and courageous exposition of one of the fundamental principles of scientific work, Kubie has pointed unerringly to the major criticism that the observational student of behavior makes of traditional experimentation in psychology, namely, its almost studied neglect of events in the natural behavior-world. How much of our so-called basic research in psychology pales beside the richness and variety of real life activity! How much is but the controlled investigation of laboratory fictions! Lazarsfeld has deplored this self-imposed limitation on psychological work in a brief, but penetrating discussion of the data of science. He writes:

> Experiments might be the only source of data for the physicist and they might be useful for the social scientist. But they are for us certainly not the main material we are working with today. Don't forget that we have data for something like 2,500 years. Plato has developed very interesting propositions on human behavior in society and, if we now come and develop models which somehow organize or clarify what Plato has to say, we are doing an important job. It isn't true that the only data in the social sciences are those which have been collected in the laboratories in the last 50 years (7, p. 99).

There is one apparent weakness in Kubie's analysis which the present writer readily acknowledges. How does one distinguish between data which require no experimental verification because they are "already manifest in nature," and data which appear to be but are not really manifest in nature? How, in other words, does the naturalist avoid the error of making "the world is flat" type of assertion?

The problem is not insoluble, although the solution is not fool-proof. We assume, first of all, that our naturalistic investigator, like the cultural anthropologist, is a well-trained student of behavior. Second, we assume that only those data concerning which a reasonable consensus can be reached will be accepted as "manifest in nature." Third, we assume that where the results of naturalistic observation contradict the already accepted body of knowledge, more complex methods of verification, not necessarily experimental in nature, will be applied. Will these safeguards eliminate all error? Certainly not. But no science has or ever will devise an error-free investigative technique. If there

is any possibility of advancing our understanding of human behavior by naturalistic observation, we should certainly be willing to bear the risks involved, recognizing, of course, that this is only one phase of the investigative operation. Hilgard expresses some appreciation of the usefulness of naturalistic study when he writes: "If we are wise enough to make the most of naturalistic observations, we may find some clues that will help us to design better experiments" (3, p. 35).

"The major contribution of experimental science," contends Kubie, "will not be limited to confirming psychoanalytic observations in the laboratory but rather to providing psychoanalysis with instruments of greater qualitative and quantitative precision." The first problem to be solved is how to record adequately the therapeutic process *"without at the same time distorting the process."* Kubie proposes an institute for research in psychoanalytic psychology, an organization which would coordinate the work of many disciplines: neurophysiology, biophysics, biochemistry, pharmacology, clinical psychology, cultural anthropology, biostatistics and, of course, psychoanalysis. He calls such an institute "a dream for the future."

In our limited judgment Kubie has correctly stated the relationship between clinical observation and controlled experimentation, and has argued convincingly for the validity of the former as a means of studying behavior. We would add only one step to his program, namely, a sociological analysis of the general and specific cultural circumstances in which the clinical knowledge is derived. Pumpian-Mindlin hints at the necessity for this when he calls attention to the fact that Freud "arose in a definite historical period and was subject to its influences." If we accept the relationist thesis of the sociology of knowledge, viz., "that it lies in the nature of certain assertions that they cannot be formulated absolutely, but only in terms of the perspective of a given situation," and if we assume that assertions about human behavior are mainly of this type, then it logically follows that the first step in determining the validity of psychological knowledge must be "an analysis based on the sociology of knowledge" (9).

Non-systematic, Reflective Observation

We come now to the third type of observational analysis, namely, non-systematic, reflective observation, as embodied in the art of the novelist or playright and in the work of the professional psychologist who prefers a non-systematic, natural behavior-world type of study. Here we are on dangerous ground. Judged by the standards of modern, experimental psychology, this approach to behavior is easily the *least* respectable of all the observational methods. Even though many psychologists would agree with Verplanck (13), that novelists and playrights "do a remarkably good job of giving plausible accounts of behavior, often in terms that seem pertinent," nevertheless, these same psychologists would probably acquiesce in the prevailing opinion that "literary psychology" is an alien rite in the House of Science. Allport's plea (1) for pluralism in the methods of psychology, as expressed in his belief that "whatever contributes to a knowledge of human nature, is an admissible method to science," seems not to have had much impact on traditional psychological thinking.

The psychologist's refusal to entertain seriously the possibility that literary creation can make a significant contribution to the understanding of human behavior rests, we believe, in what Karl Mannheim (9) would call "a one-sided and narrow orientation to the problem of knowledge." Consider the following observation, relative to the question of scientific vs. so-called pre-scientific

knowledge and understanding. It appears in Mannheim's classic work, *Ideology and Utopia*. Compare it to the typical "holier-than-thou" attitude of the experimental psychologist. We believe that the program suggested is an eminently sane and practical one, which might even have as a beneficial secondary effect, the re-establishment and strengthening of the psychologist's contact with real life behavior.

> If, however, it is true, that life affords possibilities of knowledge and understanding even where science plays no part, it is no solution to designate such knowledge as "pre-scientific" or to relegate it to the sphere of "intuition," simply in order to preserve the purity of an arbitrary definition of "science." On the contrary, it is above all our duty to inquire into the inner nature of these still unformulated types of knowledge and then to learn whether the horizons and conceptions of science cannot be so extended as to include these ostensibly pre-scientific areas of knowledge (p. 165).

The problem of validating knowledge derived from non-systematic, reflective observation can be best handled in two parts, corresponding to the two extreme instances of this form of knowledge. There is, first of all, knowledge derived from non-systematic, reflective observation that is confirmed in systematic analytical observation, clinical observation, or experimental analysis. This confirmation may be either intentional or accidental, witting or unwitting. Assuming that we can eventually agree on a set of criteria for the validity of clinically derived knowledge, such confirmation would seem to provide presumptive support for the validity of the non-systematically, reflectively derived knowledge.

One of the best examples of this manner of validating knowledge can be found in the frequent clinical confirmation of *ambivalence* as an important characteristic of affective behavior. Freud (2) attributes the coining of this term to Bleuler. Yet, but for the naming of the behavior, the ambivalent reaction was described with great insight over two thousand years ago by Lucretius (8), the celebrated Roman poet. In *Book IV* of his poem, *The Nature of the Universe*, in which Lucretius discusses sensation and sex, he writes:

> Do not think that by avoiding grand passions you are missing the delights of Venus. Rather, you are reaping such profits as carry with them no penalty. Rest assured that this pleasure is enjoyed in a purer form by the healthy than by the love-sick. Lovers' passion is storm-tossed, even in the moment of fruition, by waves of delusion and incertitude. They cannot make up their mind what to enjoy first with eye or hand. They clasp the object of their longing so tightly that the embrace is painful. They kiss so fiercely that teeth are driven into lips. All this because their pleasure is not pure, but they are goaded by an underlying impulse to hurt the thing, whatever it may be, that gives rise to these budding shoots of madness (pp. 163-164).

This example is but one of many that could be cited in which non-systematically, reflectively derived knowledge, through clinical or experimental confirmation, clearly makes a contribution to psychological understanding.

Two thousand years, we would all agree, is a long time to wait for clinical confirmation of a naturalistic observation. Surely, there must be some alterna-

tive solution to the problem of validity of non-systematic, reflective knowledge. The search for such a solution becomes imperative when we consider the second class of such knowledge, namely, non-systematically, reflectively derived knowledge which, either because it is too specific, as in the creation of a particular fictional character, or too complex, as in the fictionalized representation of an interpersonal relationship, cannot be validated in the laboratory, or in the clinic, or through systematic analytical observation. What is the status of this knowledge as seen from the scientific perspective of the psychologist? What tests of validity can be applied here?

The suggestion that we have to make is not an original one. In fact, it is its non-originality that imparts to it most of its significance. For it points to the continuing failure of modern psychology to develop methods of investigation adequate to the task of understanding real life human behavior. The suggestion is simply this: Pending the extension of the "horizons and conceptions" of psychological science so as "to include these ostensibly pre-scientific areas of knowledge," we vigorously recommend the witting, circumspect use of non-systematic, reflective knowledge as bona fide data in hypothesis and theory construction. Many psychologists, particularly those who are psychoanalytically oriented, already make use of this material, but for illustrative purposes only. What we are arguing for here is an *inductive* use of such knowledge, much as Allport has argued for an inductive use of personal documents (1). Perhaps the suggestion can be stated more simply. To wit: As long as psychologists continue to deny their birthright and leave to others—novelists, poets, playrights, social critics—the responsibility of analyzing, describing, or explaining the natural behavior-world, we had best make whatever effective use we can of their insights and observations. It would, I think, be stupidly presumptuous of us to expect the novelist or playright to adopt the methods of science as a precondition for recognizing the value of their work. The poet, the novelist, the critic, each have their own methods of arriving at understanding. If psychologists believe that the methods of science are superior ways of understanding the natural behavior-world, let them turn to this world and demonstrate the superiority of their methods. One thinks here immediately of Piaget and of the very fruitful work that he has done. But Piaget is an exception, a notable one to be sure, but an exception nonetheless.

It appears to us extremely unlikely that the insights and observations of the novelist or poet can be directly validated by experimental techniques. The complex events of the natural behavior-world are too spontaneous, have too much vitality to be faithfully recreated in the laboratory. By comparison, the prospects for clinical validation seem much brighter. Probably, the best procedure for validating naturalistic knowledge is to determine how well our understanding, prediction, and control of other observational events is improved through the use of this knowledge. Much of this material, many psychologists would agree, seems to have a high "subjective" validity. It is time we dipped into the store of riches freely offered by the poet and novelist and made effective use of it.

In this paper we have argued for a pluralistic approach to the understanding of human behavior. Underlying this plea is a genuine fear that the psychologist's obsessive concern with scientific respectability may blind him to the avenues of understanding opened up by a non-experimental, observational type analysis. We agree with Krutch—although probably had we had the happy fortune to have first formulated this criticism, we would have spoken more circumspectly—that:

Perhaps Hamlet was nearer right than Pavlov. Perhaps the exclamation "How like a god!" is actually more appropriate than "How like a dog! How like a rat! How like a machine!" Perhaps we have been deluded by the fact that the methods employed for the study of man have been for the most part those originally devised for the study of machines or the study of rats, and are capable, therefore, of detecting and measuring only those characteristics which the three do have in common. But we have already gone a long way on the contrary assumption, and we take it more completely for granted than we sometimes realize. The road back is not an easy one (5, pp. 32-33).

REFERENCES

1. ALLPORT, G. W. *The use of personal documents in psychological science.* New York: Social Science Research Council, 1942.
2. FREUD, S. *The basic writings of Sigmund Freud.* New York: Random House, 1938. (English translation by A. A. Brill.)
3. HILGARD, E. R. Experimental approaches to psychoanalysis. In E. Pumpian-Mindlin (Ed.), *Psychoanalysis as science.* Stanford, California: Stanford University Press, 1952. Pp. 3-45.
4. KRUTCH, J. W. *The modern temper.* New York: Harcourt, Brace, 1929 (Harvest edition).
5. KRUTCH, J. W. *The measure of man.* New York: Grosset & Dunlap, 1953.
6. KUBIE, L. S. Problems and techniques of psychoanalytic validation and progress. In E. Pumpian-Mindlin (Ed.), *Psychoanalysis as science.* Stanford, California: Stanford University Press, 1952. Pp. 46-124.
7. LAZARSFELD, P. Concluding remarks. In *Mathematical models of human behavior.* Stamford, Connecticut: Dunlap, 1955. Pp. 97-103.
8. LUCRETIUS. *The nature of the universe.* Harmondsworth, England: Penguin, 1951. (English translation by R. E. Latham.)
9. MANNHEIM, K. *Ideology and utopia.* New York: Harcourt, Brace, 1936 (Harvest edition).
10. PUMPIAN-MINDLIN, E. The position of psychoanalysis in relation to the biological and social sciences. In E. Pumpian-Mindlin (Ed.), *Psychoanalysis as science.* Stanford, California: Stanford University Press, 1952. Pp. 125-158.
11. PUMPIAN-MINDLIN, E. (Ed.) *Psychoanalysis as science.* Stanford, California: Stanford University Press, 1952.
12. SWARTZ, P. On the validity of the experimental approach to behavior. *Psychol. Rec.*, 1957, 7, 119-122.
13. VERPLANCK, W. S. The operant, from rat to man: an introduction to some recent experiments on human behavior. *Trans. N. Y. Acad. Sci.*, 1955, 17, Ser. II, 594-601.